

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского»

УТВЕРЖДЕНО
решением ученого совета ННГУ
протокол от "30" ноября 2022 г. № 13

Рабочая программа дисциплины
Прикладная статистика в среде R

Уровень высшего образования
Подготовка научных и научно-педагогических кадров

Программа аспирантуры
Экология

Научная специальность
1.5.15 Экология

Форма обучения
Очная

Нижний Новгород
2023 год

1. Место и цель дисциплины в структуре ОПОП

Дисциплина «Прикладная статистика в среде R» относится к числу *элективных* дисциплин образовательного компонента программы аспирантуры и изучается на 2 году обучения в 3 семестре.

Цель дисциплины – *овладение (актуализация) навыками работы в среде R, освоение методов импорта/экспорта и манипулирования данными, изучение современных методов одномерного и многомерного статистического анализа, изучение современных средств визуализации данных и результатов анализа, овладение практическими навыками анализа и визуализации данных в R.*

2. Планируемые результаты обучения по дисциплине

Выпускник, освоивший программу, должен

Знать: основные проблемы биологии и основные теории экологии, методы генерирования новых идей при решении фундаментальных и прикладных проблем в области экологии, современные методы исследования в области экологии и принципы работы исследовательской аппаратуры, основные методы проведения экспертизы, способы и методы внедрения результатов научно-исследовательских работ в практику

Уметь: осуществлять поиск, анализ и обобщение современных научных данных, в том числе с использованием международных систем индексирования; ставить задачу, планировать и выполнять научные исследования для разработки принципов и механизмов, обеспечивающих устойчивое развитие при сохранении стабильного состояния природной среды с использованием современной аппаратуры, вычислительных средств, компьютерных технологий, анализировать варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации научно-исследовательских работ в практику.

Владеть: навыками использования современных методов исследования и информационно-коммуникационных технологий

3. Структура и содержание дисциплины.

Объем дисциплины (модуля) составляет 3 з.е., всего - 108 часа, из которых 36 часов составляет контактная работа обучающегося с преподавателем (занятия семинарского типа – 36 часов), 72 часа составляет самостоятельная работа обучающегося.

Таблица 2

Структура дисциплины

Наименование раздела дисциплины	Всего, часов	В том числе					
		Контактная работа, часов					Самостоятельная работа обучающегося, часов
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Консультации	Всего	
1. Основы работы в R и RStudio	13		4			4	9
2. Работа с данными в R	34		12			12	22
3. Основы графической системы R	14		4			4	10
4. Работа с распределениями вероятностей и описательная статистика	14		4			4	10
5. Методы одномерного статистического анализа	19		8			8	11
6. Методы многомерного статистического анализа	14		4			4	10

Промежуточная аттестация: –	Зачет					
Итого	108		36		36	72

Таблица 3

Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела	Форма проведения занятия	Форма текущего контроля*
1.	Основы работы в R и RStudio	Установка R и RStudio. Основы интерфейса RStudio. Типы данных (числовые, логические, строковые, специальные). Приведение типов.	научно-практическое занятие	Тест
2.	Работа с данными в R	Структуры данных (вектора, матрицы, массивы, списки, фреймы, факторы). Обращение к данным (индексирование векторов, матриц и списков). Управляющие структуры (условный оператор, циклы). Функции (определение, аргументы, возвращаемое значение). Векторизованные вычисления. Манипулирование данными (создание и удаление переменных, выборки, ранжирование).	научно-практическое занятие	Тест, расчетно-графическое задание
3.	Основы графической системы R	Графические подсистемы. Функция plot(). Графические параметры. Графические устройства.	научно-практическое занятие	Тест, расчетно-графическое задание
4	Работа с распределениями вероятностей и описательная статистика	Распределения вероятностей и работа с ними в R. Описание и визуализация распределения количественных данных. Описание и визуализация распределения качественных данных. Нормальное распределение и диагностика отклонений.	научно-практическое занятие	Тест, расчетно-графическое задание
5.	Методы одномерного статистического анализа	Простейшие гипотезы: одно- и двухвыборочные сравнения. Сравнение нескольких групп: однофакторный дисперсионный анализ и критерий Крускала-Уоллиса. Множественные сравнения. Простая регрессия и корреляция. Статистика нечисловых данных.	научно-практическое занятие	Тест, расчетно-графическое задание
6.	Методы многомерного статистического анализа	Анализ комплекса признаков. Многомерный дисперсионный анализ. Анализ главных компонент. Линейный дискриминантный анализ.	научно-практическое занятие	Тест, расчетно-графическое задание

4. Формы организации и контроля самостоятельной работы обучающихся

Виды самостоятельной работы аспирантов в рамках освоения дисциплины:

- изучение понятийного аппарата и проработка тем дисциплины;
- работа с основной и дополнительной литературой дома и в библиотеке;
- изучение сайтов по темам дисциплины в сети Интернет;
- подготовка к тестам (примеры заданий см. в п. 6.3);
- подготовка к решению задач (примеры заданий см. в п. 6.3);
- подготовка к экзамену.

Изучение понятийного аппарата дисциплины

Вся система индивидуальной самостоятельной работы должна быть подчинена усвоению понятийного аппарата, поскольку одной из важнейших задач подготовки современного грамотного специалиста является овладение и грамотное применение профессиональной терминологии. Лучшему усвоению и пониманию дисциплины помогут различные энциклопедии, словари, справочники и другие материалы, указанные списке литературы.

Работа над основной и дополнительной литературой

Изучение рекомендованной литературы следует начинать с учебников и учебных пособий, затем переходить к нормативно-правовым актам, научным монографиям и материалам периодических изданий. Конспектирование – одна из основных форм самостоятельного труда, требующая от студента активно работать с учебной литературой и не ограничиваться конспектом лекций.

Аспирант должен уметь самостоятельно подбирать необходимую для учебной и научной работы литературу. При этом следует обращаться к предметным каталогам и библиографическим справочникам, которые имеются в библиотеках. Для аккумуляции информации по изучаемым темам рекомендуется формировать личный архив, а также каталог используемых источников.

Изучение сайтов по темам дисциплины в сети Интернет

Ресурсы Интернет являются одним из альтернативных источников быстрого поиска требуемой информации. Их использование возможно для получения основных и дополнительных сведений по изучаемым материалам.

Самостоятельная работа при подготовке к экзамену

Контроль выступает формой обратной связи и предусматривает оценку успеваемости аспирантов и разработку мер по дальнейшему повышению качества подготовки современных специалистов.

В начале семестра рекомендуется внимательно изучить перечень вопросов к зачету по данной дисциплине, а также использовать в процессе обучения программу, другие методические материалы, разработанные кафедрой по данной дисциплине. Это позволит в процессе изучения тем сформировать более правильное и обобщенное видение студентом существа того или иного вопроса за счет:

- а) уточняющих вопросов преподавателю;
- б) самостоятельного уточнения вопросов на смежных дисциплинах;
- в) углубленного изучения вопросов темы по учебным пособиям.

После изучения соответствующей тематики рекомендуется проверить наличие и формулировки вопроса по этой теме в перечне вопросов к экзамену, а также попытаться изложить ответ на этот вопрос. Если возникают сложности при раскрытии материала, следует вновь обратиться к лекционному материалу, материалам практических занятий, уточнить терминологический аппарат темы, а также проконсультироваться с преподавателем.

5. Фонд оценочных средств для аттестации по дисциплине

5.1. Критерии и процедуры оценивания результатов обучения по дисциплине.

При выполнении всех работ учитываются следующие **основные критерии**:

- уровень теоретических знаний (подразумевается не только формальное воспроизведение информации, но и понимание предмета, которое подтверждается правильными ответами на дополнительные, уточняющие вопросы, заданные членами комиссии);
- умение использовать теоретические знания при анализе конкретных проблем, ситуаций;
- качество изложения материала, то есть обоснованность, четкость, логичность ответа, а также его полнота (то есть содержательность, не исключающая сжатости);
- способность устанавливать внутри- и межпредметные связи,
- оригинальность мышления, знакомство с дополнительной литературой и другие факторы.

Описание шкалы оценивания на промежуточной аттестации в форме зачета

Оценка	Уровень подготовленности, характеризуемый оценкой
<i>Зачтено</i>	владение программным материалом, понимание сущности рассматриваемых процессов и явлений, умение самостоятельно обозначить проблемные ситуации в организации научных исследований, способность критически анализировать и сравнивать существующие подходы и методы к оценке результативности научной деятельности, свободное владение источниками, умение четко и ясно излагать результаты собственной работы, следовать нормам, принятым в научных дискуссиях.
<i>Не зачтено</i>	непонимание смысла ключевых проблем, недостаточное владение науковедческой терминологией, неумение самостоятельно обозначить проблемные ситуации, неспособность анализировать и сравнивать существующие концепции, подходы и методы, неумение ясно излагать результаты собственной работы, следовать нормам, принятым в научных дискуссиях.

Критерии оценивания тестов

Тестовые задания оцениваются по пятибалльной системе в зависимости от доли правильных ответов или правильно выполненных контрольных заданий:

- «отлично»: 80–100% правильных ответов;
- «хорошо»: 65–80% правильных ответов;
- «удовлетворительно»: 50–65% правильных ответов;
- «неудовлетворительно» – 25–50% правильных ответов;
- «плохо» – менее 25% правильных ответов.

Критерии оценивания расчетно-графических заданий

Расчетно-графические задания состоят из нескольких задач, подразумевающих определенную последовательность действий. Расчетно-графические задания оцениваются по альтернативной шкале в зависимости от доли верно выполненных задач:

- «зачтено»: 50–100% правильно выполненных задач;
- «не зачтено»: менее 50% правильно выполненных задач.

5.2. Примеры типовых контрольных заданий или иных материалов, используемых для оценивания результатов обучения по дисциплине

Вопросы к экзамену

1. Установка R и RStudio. Основы интерфейса RStudio.
2. Типы данных (числовые, логические, строковые, специальные). Приведение типов.
3. Структуры данных (вектора, матрицы, массивы, списки, фреймы, факторы).
4. Обращение к данным (индексирование векторов, матриц и списков).
5. Управляющие структуры (условный оператор, циклы).
6. Функции (определение, аргументы, возвращаемое значение).
7. Векторизованные вычисления.
8. Манипулирование данными (создание и удаление переменных, выборки, ранжирование).
9. Графические подсистемы в R.
10. Функция plot().
11. Графические параметры.
12. Графические устройства.
13. Работа с распределениями вероятностей в R.
14. Описание и визуализация распределения количественных данных.
15. Описание и визуализация распределения качественных данных.
16. Нормальное распределение и диагностика отклонений.
17. Простейшие гипотезы: одно- и двухвыборочные сравнения.
18. Сравнение нескольких групп: однофакторный дисперсионный анализ и критерий Крускала-Уоллиса.
19. Множественные сравнения.
20. Простая регрессия и корреляция.
21. Анализ взаимосвязи качественных переменных.
22. Соответствие распределения качественной переменной теоретическому распределению.
23. Анализ комплекса признаков.
24. Многомерный дисперсионный анализ.
25. Анализ главных компонент.
26. Линейный дискриминантный анализ.

Типовые тестовые задания:

1. R – это свободно распространяемая версия: а) языка C; б) языка S; в) программы Statistica; г) среды вычислений Matlab.
2. Символом начала комментария в скриптах R служит: а) #; б) %; в) &; г) \$.
3. Какое сочетание клавиш служит для выполнения блока кода, выделенного в редакторе скрипта? а) Ctrl + S; б) Alt + R; в) Shift + Enter; г) Ctrl + R.
4. Что произойдет при попытке обращения к элементу вектора, номер которого превышает длину вектора? а) ошибка; б) будет возвращено значение NA; в) будет возвращена длина вектора; г) будет возвращен 0.
5. Результат выражения rep(1:3, times = 3)[6] : а) 1; б) 2; в) 3; г) NA.

6. Оператор обращения к элементам вектора: а) (); б) []; в) {}; г) [()].

7. Какой аргумент используется для указания разделителя целой и дробной частей при загрузке текстовых файлов? а) header; б) nrow; в) dec; г) sep.

8. Выберите верное утверждение : а) пакеты пишутся только разработчиками R; б) все пакеты хранятся на сайте CRAN; в) пакеты – это совокупности функций и наборов данных; г) пакеты необходимо устанавливать каждый раз перед использованием.

9. Для чего служит графический параметр mfrow ? а) для задания размера графика; б) для создания многопанельных графиков; в) для определения типа легенды; г) для контроля ширины полей вокруг графика.

10. Что не может являться графическим устройством? а) окно; б) текстовый файл; в) файл формата pdf; г) файл формата png.

Типовые тестовые задания:

1. Изменчивость выборочной статистики заключается в том, что:

- а) значение выборочной статистики отличается от генерального параметра;
- б) значения выборочной статистики отличаются для выборок разного объема;
- в) разные выборочные статистики имеют разные значения;
- г) значения выборочной статистики неодинаковы для разных выборок одного объема.

2. Выберите неверное утверждение относительно среднего распределения выборочных средних:

- а) оно зависит от объема выборки;
- б) оно равно среднему для всех возможных выборок;
- в) оно равно генеральному среднему;
- г) его точное значение на практике установить нельзя.

3. Из генеральной совокупности со средним 40 и стандартным отклонением 7.5 берутся выборки объемом 25 наблюдений. Чему равно среднее распределения выборочных средних?

- а) 7.5; б) 8; в) 40; г) 5.

4. Из генеральной совокупности со средним 40 и стандартным отклонением 7.5 берутся выборки объемом 25 наблюдений. Чему равна стандартная ошибка?

- а) 7.5; б) 0.3; в) 8; г) 1.5.

5. При использовании критерия Шапиро-Уилка получено р-значение 0.03. Это значит, что:

- а) выборка соответствует нормальному распределению;
- б) нуль-гипотезу о характере распределения выборки, отвергнуть нельзя;

в) выборка отклоняется от нормального распределения на 5%-ном уровне значимости;

г) среднее генеральной совокупности не отличается от нуля.

Типовые расчетно-графические задания:

Расчетно-графическая работа 1:

1. Создайте следующий вектор n1 без использования функции конкатенации c():

(1, 1.7, 2.4, 3.1, 3.8, 4.5, 5.2, 5.9, 6.6)

2. Создайте следующий вектор n2 с помощью функции rep():

(-1, 8, 9, -1, 8, 9, -1, 8, 9, -1, 8)

3. Вычислите отношение суммы элементов 5 и 7 вектора n1 к сумме элементов 5 и 11 вектора n2.

4. Вычислите квадратный корень от произведения длин векторов n1 и n2.

5. Объедините в вектор n3 вектор n1 и все элементы вектора n2 за исключением 3-го и 11-го.

6. Замените элемент 17 вектора n3 натуральным логарифмом элемента 10 вектора n2.

7. Создайте следующий логический вектор index:

(TRUE, FALSE, TRUE, FALSE, TRUE, FALSE, TRUE, FALSE, TRUE)

8. Выберите четные элементы вектора n1 с помощью логического вектора index.

9. Выберите из вектора n3 все элементы, которые не равны 4.15.

10. Выберите из вектора n2 все отрицательные элементы.

11. Создайте текстовый вектор роет из первой строфы стихотворения 'Зимнее утро' А.С. Пушкина таким образом, чтобы каждая строка соответствовала одному элементу.

12. Вычислите число символов третьей строки из вектора роет.

13. Создайте вектор letters, составленный из первых букв каждой из строк вектора роет.

14. Сохраните вектора n1, n2, n3 и роет в файл surname.hw2.rda.

15. Сохраните итоговый скрипт в файл surname.hw2.r.

16. Отправьте два файла на адрес преподавателя в сообщении с темой group.surname.hw2.

Типовое расчетно-графическое задание:

Расчетно-графическая работа 2:

Описание данных. В файле benthos.xls содержатся данные о видовой структуре макрозообентоса р. Кудьма, собранные на 16 станциях (переменная site) в 2012-2013 гг (переменная dt). Данные представляют собой значения биомассы (bm), численности (den) и видового богатства (S) макрозообентоса, а также значения индексов разнообразия Шеннона (H) и Симпсона (C) и индекса доминирования Симпсона (Ds).

1. Загрузите основной блок данных во фрейм bent программным способом (не используя буфер обмена).

2. Выведите в консоль размерность импортированного фрейма.

3. Вычислите, сколько наблюдений фрейма bent содержат полные данные (в соответствующих строках нет значений NA).

4. Рассчитайте среднее численности зообентоса.
5. Выберите в новый фрейм bent1 данные за октябрь 2013 г.

Указание: создайте логический вектор с использованием оператора сравнения и используйте его для индексирования исходного фрейма.

6. Рассчитайте среднее индекса разнообразия Шеннона по данным за выбранную дату.
7. Выберите в новый фрейм bent2 данные по 5 станции.
8. Рассчитайте, сколько раз в данных по выбранной станции встречается минимальное значение видового богатства.

Указание: рассчитайте минимум, используйте оператор сравнения и суммирование полученного логического вектора.

9. Рассчитайте разность средних значений индекса разнообразия Симпсона станций 4 и 11.

6. Учебно-методическое и информационное обеспечение дисциплины.

а) Основная литература

1. Шипунов А.Б., Балдин Е.М., Волкова П.А., Коробейников А.И., Назарова С.А., Петров С.В., Суфиянов В.Г. Наглядная статистика. Используем R! М.: ДМК Пресс, 2012. 298 с. – Доступно на ЭБС «Консультант студента». Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785940748281.html>

2. Зорин А.В., Федоткин М.А. Введение в прикладной статистический анализ в пакете R. Учебно-методическое пособие. Н. Новгород: ННГУ, 2010. 50 с. – Зарегистрирован в фонде образовательных электронных ресурсов (компьютерных изданий) ННГУ. Режим доступа: <http://www.unn.ru/pages/e-library/methodmaterial/2010/3.pdf>

3. Наац В.И., Наац И.Э. Математические модели и численные методы в задачах экологического мониторинга атмосферы – М.: ФИЗМАТЛИТ, 2009. – 327 с. – Доступно на ЭБС «Консультант студента». Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785922111607.html>

б) Дополнительная литература

1. Трухачёва Н.В. Математическая статистика в медико-биологических исследованиях с применением пакета Statistica – М.: ГЭОТАР-Медиа, 2013. – 384 с. – Доступно на ЭБС «Консультант студента». Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785970425671.html>

2. Николайкин Н.И., Николайкина Н.Е., Мелехова О.П. Экология – М.: Дрофа, 2006. – (1 экз. в библиотеке ННГУ)

в) Программное обеспечение и Интернет-ресурсы

1. <http://www.r-project.org/> – R Project: The R Project for Statistical Computing – язык программирования и среда разработки для статистических вычислений.

2. <http://cran.r-project.org/> – CRAN: The Comprehensive R Archive – архив пакетов для расширения языка программирования R.

3. <http://www.rstudio.com/> – R Studio – среда разработки для языка программирования R.
4. <http://www.statsoft.ru/home/textbook/> – электронный учебник по статистике и планированию эксперимента.
5. <http://r-analytics.blogspot.ru/> – R: Анализ и визуализация данных.

7. Материально-техническое обеспечение дисциплины

- помещения для проведения занятий: лекционного типа, семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для хранения и профилактического обслуживания оборудования и помещения для самостоятельной работы обучающихся, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду ННГУ;
 - материально-техническое обеспечение, необходимое для реализации дисциплины, включая лабораторное оборудование;
 - лицензионное программное обеспечение: *Windows, Microsoft Office*;
 - обучающиеся из числа лиц с ограниченными возможностями здоровья обеспечиваются электронными и (или) печатными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.
- ресурсам.

Рабочая программа учебной дисциплины составлена в соответствии с учебным планом, Положением о подготовке научных и научно-педагогических кадров в аспирантуре (адъюнктуре) (Постановление Правительства РФ от 30.11.2021 № 2122), Федеральными государственными требованиями к структуре программ подготовки научных и научно-педагогических кадров в аспирантуре (адъюнктуре) (Приказ Минобрнауки РФ от 20.10.2021 № 951).

Авторы:

Авторы Якимов В.Н.

Рецензент(ы) Кравченко Г.А.

Заведующий кафедрой Якимов В.Н.

Программа одобрена на заседании Методической комиссии Института биологии и биомедицины от 06.09.2022 года, протокол № 1.